Appendix A: Maps

Figures 7, 8 & 9 show the full sets of BK15 $T/Q/U$ maps at 95, 150 & 220 GHz. The right side of each figure shows realizations of noise created by randomly flipping the sign of data subsets while coadding the map—see Sec. V.B of Ref. [1] for further details.

Appendix B: 95 GHz and 220 GHz Internal Consistency Tests

A powerful internal consistency test are data split difference tests which we refer to as "jackknives". As well as the full coadd signal maps we also form many pairs of split maps where the splits are chosen such that one might expect different systematic contamination in the two halves of the split. The split halves are differenced and the power spectra taken. We then take the deviations of these from the mean of signal+noise simulations and form χ^2 and χ (sum of deviations) statistics. In this section we perform tests of the 95 GHz and 220 GHz data sets which are exactly analogous to the tests of the 150 GHz data sets performed in Sec. VII.C of Ref. [1] and Sec. 6.3 of Ref. [3]. (Since going from 8 to 9 receiver-years of *Keck* 150 GHz barely shifts the results we omit those tests for brevity.)

Tables I and II show the χ^2 and χ statistics for the 95 GHz and 220 GHz jackknife tests respectively, while Figures 10 & 11 present the same results in graphical form. Note that these values are partially correlated—particularly the 1–5 and 1–9 versions of each statistic. We conclude that there is no evidence for corruption of the data at a level exceeding the noise.

Appendix C: 95 GHz Spectral Stability

We next test the mutual compatibility of the 2014 and 2015 95 GHz spectra. We compare the differences of the real spectra to the differences of simulations which share the same underlying input skies. We perform the test in two ways: firstly by differencing the single season spectra ($K2014_{95}$ and $K2015_{95}$), and secondly by differencing the 2014 single season from the 2014+2015 season combined spectrum. Fig. 12 shows the results—the differences are seen to be consistent with noise fluctuation.

Appendix D: Additional Spectra

Fig. 2 shows only a small subset of the spectra which are used in the likelihood analysis and included in the `COSMOMC` input file. We are using three BICEP2/*Keck* bands, two WMAP bands, and seven *Planck* bands resulting in 12 auto- and 66 cross-spectra. In Fig. 13 we show all of these together with the maximum likelihood model from the baseline analysis whose parameters were

TABLE I. Jackknife PTE values from χ^2 and χ (sum of deviations) tests for *Keck Array* 95 GHz data taken in 2014 and 2015. This table is analogous to Table I of Ref. [4] but extended to two seasons of data.

Jackknife	Band powers 1–5 χ^2	Band powers 1–9 χ^2	Band powers 1–5 χ	Band powers 1–9 χ
Deck jackknife				
EE	0.042	0.176	0.421	0.501
BB	0.132	0.186	0.852	0.952
EB	0.705	0.922	0.196	0.361
Scan Dir jackknife				
EE	0.281	0.136	0.553	0.920
BB	0.154	0.100	0.980	0.968
EB	0.269	0.263	0.096	0.050
Tag Split jackknife				
EE	0.194	0.377	0.743	0.930
BB	0.084	0.160	0.920	0.898
EB	0.685	0.870	0.259	0.319
Tile jackknife				
EE	0.321	0.517	0.800	0.916
BB	0.862	0.978	0.832	0.792
EB	0.363	0.279	0.758	0.711
Phase jackknife				
EE	0.858	0.800	0.627	0.621
BB	0.010	0.048	0.186	0.200
EB	0.337	0.423	0.721	0.758
Mux Col jackknife				
EE	0.778	0.912	0.904	0.804
BB	0.651	0.497	0.419	0.880
EB	0.343	0.224	0.569	0.253
Alt Deck jackknife				
EE	0.110	0.409	0.399	0.483
BB	0.335	0.487	0.569	0.677
EB	0.643	0.347	0.517	0.950
Mux Row jackknife				
EE	0.459	0.557	0.599	0.896
BB	0.784	0.447	0.665	0.832
EB	0.697	0.621	0.132	0.042
Tile/Deck jackknife				
EE	0.393	0.693	0.812	0.691
BB	0.267	0.309	0.303	0.333
EB	0.579	0.355	0.760	0.934
Focal Plane inner/outer jackknife				
EE	0.617	0.419	0.906	0.992
BB	0.132	0.226	0.892	0.972
EB	0.984	0.629	0.683	0.806
Tile top/bottom jackknife				
EE	0.595	0.020	0.593	0.407
BB	0.954	0.990	0.615	0.357
EB	0.289	0.505	0.954	0.840
Tile inner/outer jackknife				
EE	0.305	0.605	0.158	0.090
BB	0.509	0.601	0.527	0.567
EB	0.449	0.447	0.375	0.096
Moon jackknife				
EE	0.086	0.299	0.066	0.086
BB	0.900	0.852	0.291	0.325
EB	0.200	0.477	0.782	0.796
A/B offset best/worst				
EE	0.090	0.034	0.766	0.295
BB	0.882	0.435	0.806	0.970
EB	0.613	0.902	0.611	0.561

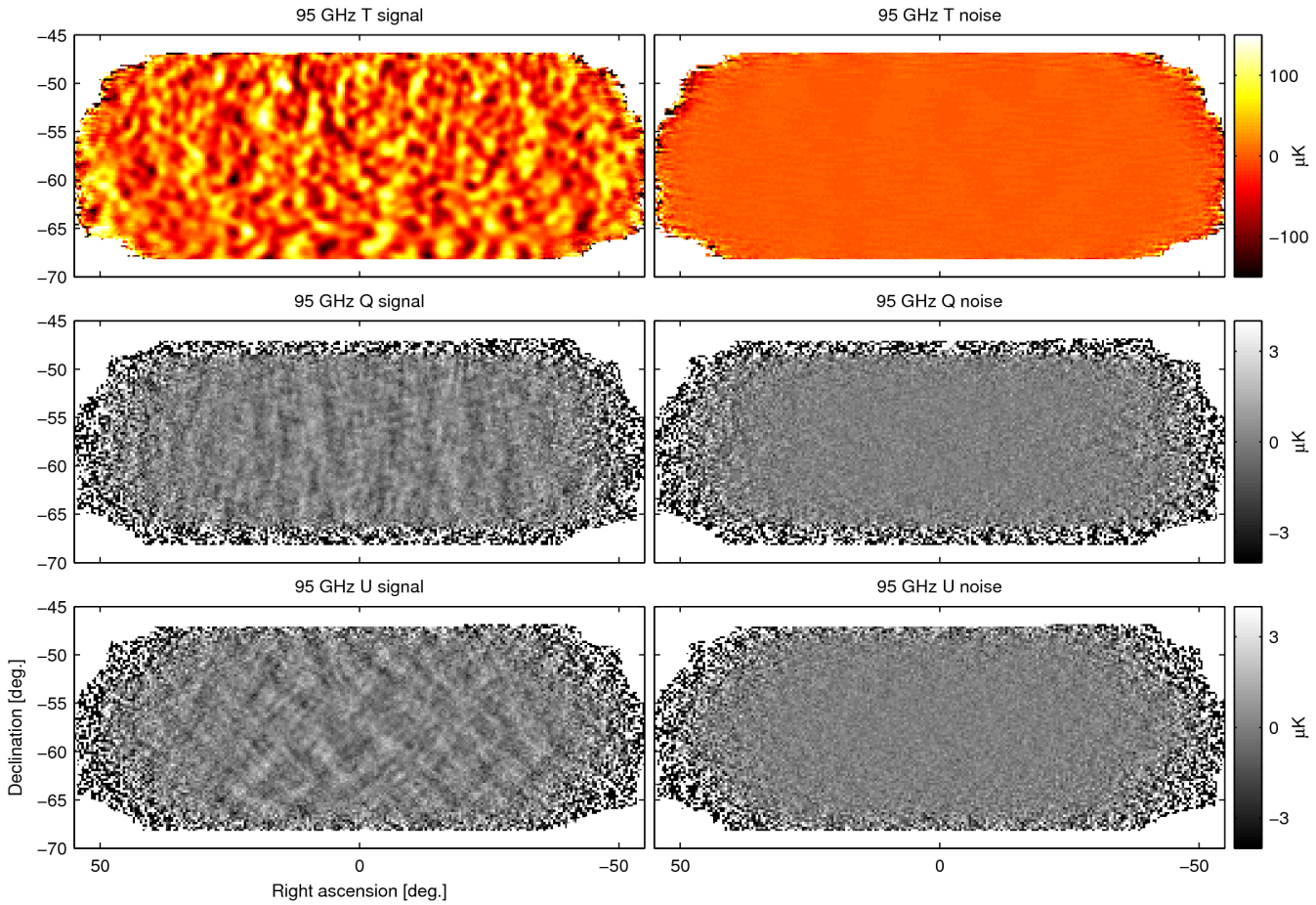

FIG. 7. T, Q, U maps at 95 GHz using data taken by two receivers of *Keck Array* during the 2014 & 2015 seasons—we refer to these maps as $\mathrm{BK15}_{95}$. The left column shows the real data maps with 0.25° pixelization as output by the reduction pipeline. The right column shows a noise realization made by randomly assigning positive and negative signs while coadding the data. These maps are filtered by the instrument beam (FWHM 43 arcmin [2]), timestream processing, and (for Q & U) deprojection of beam systematics. Note that the horizontal/vertical and 45° structures seen in the Q and U signal maps are expected for an E-mode dominated sky.

quoted above. Most of the spectra not already shown in Fig. 2 have low signal-to-noise, although a few of them carry interesting additional information on the possible level of synchrotron as will be noted later.

The HL likelihood [5] we use for the primary analysis accounts for the full joint PDF of auto- and cross-spectral bandpowers which are derived from maps which are a combination of (correlated) signal and (mostly uncorrelated) noise. We choose to quantify the absolute goodness-of-fit of the data to the maximum likelihood model using a simple χ^2 statistic which assumes that the bandpowers are normally distributed about their expectation values. We find that the distribution of this χ^2 statistic for the standard (499) lensed-ΛCDM+dust+noise simulations versus their input model is significantly broader than the nominal theoretical distribution—presumably because of the non-normal distribution of the bandpowers. It is therefore most appropriate to compare the real data value to the simulated distribution.

For the $9\times78 = 702$ bandpowers shown in Fig. 13, $\chi^2 = (d-m)^T C^{-1}(d-m) = 760$, where d are the bandpower values, m are the model expectation values, and C is the bandpower covariance matrix. This has a nominal theory PTE of 0.06 but a PTE versus the simulations of 0.19. If instead we take the sum of the normalized deviations $(\sum((d-m)/e)^2$ where e is the square-root of the diagonal of C) we find that the PTE versus the simulations is 0.23. We conclude that the parametric model which we have chosen—in combination with the approximation of Gaussian fluctuation of the dust (and synchrotron) sky patterns—is an adequate description of the presently available data.

We also run a likelihood analysis to find the CMB and foreground contributions on a bandpower-by-bandpower basis. The baseline analysis is a single fit to all 9 bandpowers across 78 spectra with 8 parameters. Instead we now perform 9 separate fits—one for each bandpower—

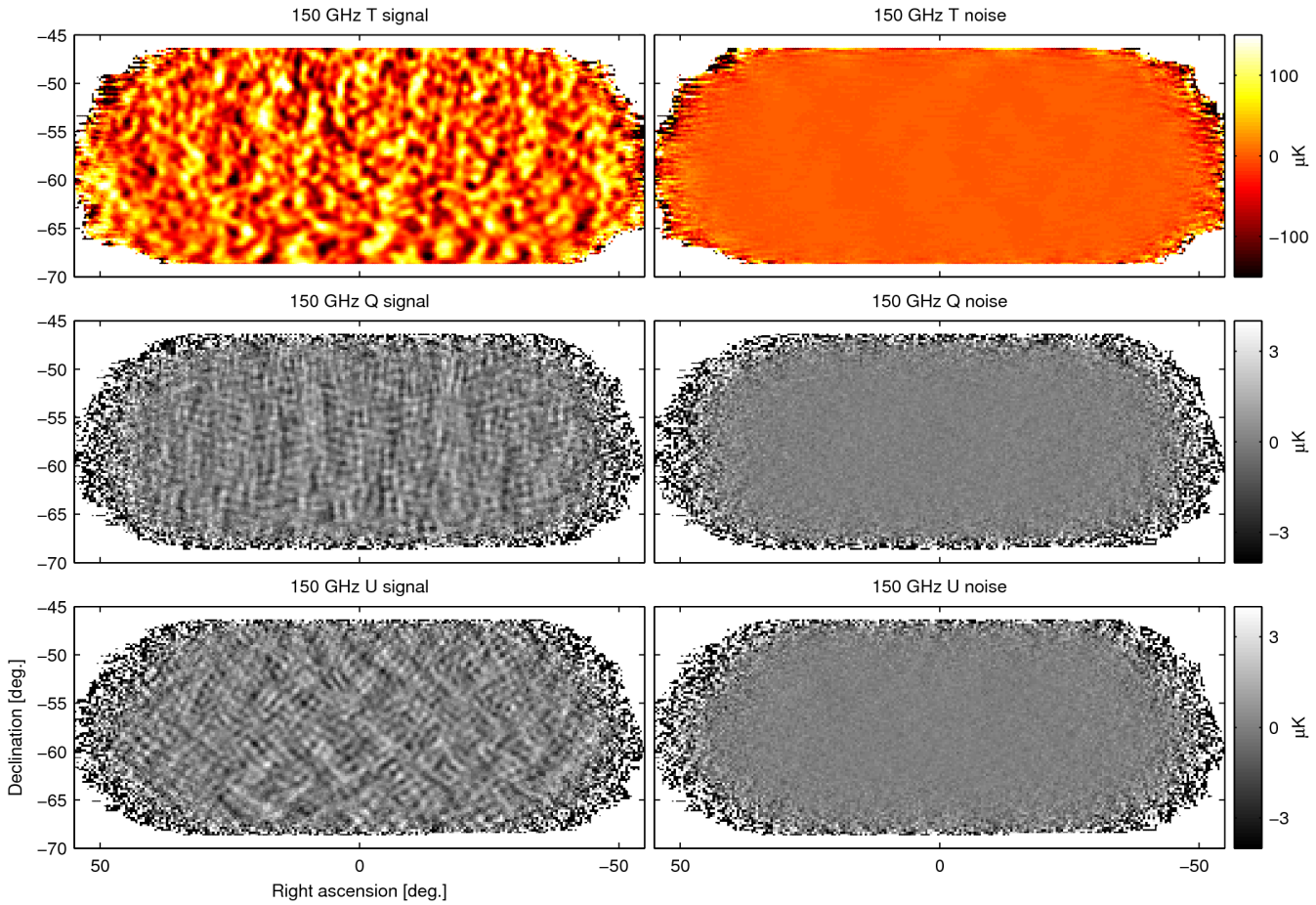

FIG. 8. T, Q, U maps at 150 GHz using all BICEP2/*Keck* data up to and including that taken during the 2015 observing season—we refer to these maps as $\mathrm{BK15}_{150}$. These maps are directly analogous to the 95 GHz maps shown in Fig. 7 except that the instrument beam filtering is in this case 30 arcmin FWHM [2].

across the 78 spectra, with 6 parameters in each fit. These 6 parameters are the amplitudes of CMB, dust and synchrotron plus β_d, β_s and ϵ with identical priors to the baseline analysis. The results are shown in Fig. 14—the resulting CMB values are consistent with lensed-ΛCDM while the dust values are consistent with the level of dust found in the baseline analysis. Synchrotron is tightly limited in all the multipole ranges, and not detected in any of them.

Appendix E: Likelihood Variation and Validation

1. Likelihood Evolution

We make only one model change versus the BK14 baseline analysis—we extend the range over which the sync/dust correlation parameter is marginalized from $0 < \epsilon < 1$ to the full possible range $-1 < \epsilon < 1$. This change was motivated by noting that the likelihood of this parameter peaked at zero in the BK14 analysis and following the philosophy of "allowing the data to select the model it prefers so long as this does not result in bias on r." While we are not aware of any theoretical motivation to consider negative values, anti-correlation is presumably physically possible. Empirical evidence is sparse; Ref. [6] reports a correlation of 0.2 for $30 < \ell < 200$, but the most recent *Planck* analysis detects (positive) sync/dust correlation only for $\ell < 50$ [7].

Fig. 15 shows the sequence of steps from the BK14 baseline analysis to the new baseline. Changing the ϵ marginalization range results in the change from green to magenta. Adding the 2015 data at 95 & 150 GHz causes the change from magenta to blue. Finally adding the new 220 GHz band results in the change from blue to black. The net result is a narrowing of the r likelihood curve and a slight downward shift in the peak position. Note that we made the choice to change the ϵ prior based on the considerations above, and before looking at these real data results.

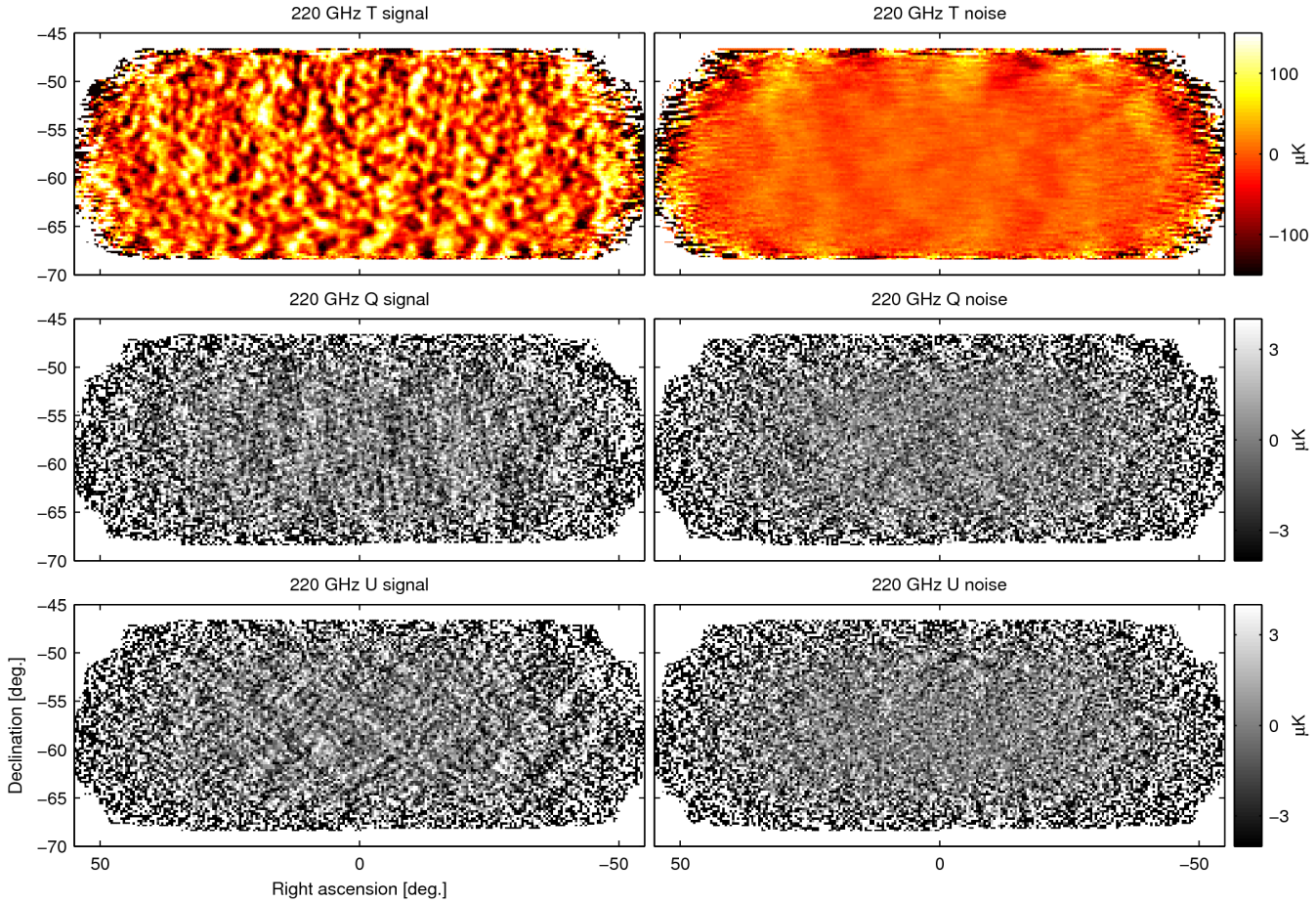

FIG. 9. T, Q, U maps at 220 GHz using data taken by two receivers of *Keck Array* during the 2015 season—we refer to these maps as BK15$_{220}$. These maps are directly analogous to the 95 GHz maps shown in Fig. 7 except that the instrument beam filtering is in this case 20 arcmin FWHM [2].

2. Likelihood Variation

Fig. 16 shows some variations from the baseline analysis choices. The HL likelihood [5] requires that one provide a "fiducial model", but it is not supposed to matter very much what this model is so long as it is reasonably close to reality. Since the BKP paper we have used $A_{\rm d,353} = 3.6\,\mu{\rm K}^2$, $A_{\rm sync} = 0$, $r = 0$. Switching to $A_{\rm d,353} = 5\,\mu{\rm K}^2$, $A_{\rm sync} = 0$, $r = 0.05$ (blue) or $A_{\rm d,353} = 5\,\mu{\rm K}^2$, $A_{\rm sync,23} = 2\,\mu{\rm K}^2$, $r = 0.05$ (red) makes little difference.

Since BKP our baseline analysis has used a prior on the frequency spectral index of dust of $\beta_{\rm d} = 1.59\pm0.11$, using a Gaussian prior with the given 1σ width. These numbers are based on external information from *Planck* [8, 9] derived from other regions of the sky. In BK14 removing this prior resulted in a significant upshift in the r constraint curve and a shift and broadening of the $A_{\rm d}$ curve. However, with the addition of the new *Keck* 220 GHz data we are now able to constrain $\beta_{\rm d}$ sufficiently well that changes when removing this prior are small (black to magenta). The $\beta_{\rm d}$ constraint curve (not shown) is close to Gaussian in shape with mean/σ of 1.65/0.20. With further improvements in the data in the future we will no longer need the $\beta_{\rm d}$ prior and hence will be able to remove the uncertainty that comes from assuming that dust behavior in our sky patch is the same as the average behavior over larger regions of sky.

Our baseline prior on the frequency spectral index of synchrotron is $\beta_{\rm s} = -3.1\pm0.3$ [10], with a Gaussian shape with the given 1σ width. Relaxing to a uniform prior over the range $-4.5 < \beta_{\rm s} < -2.0$ produces no significant changes (black to green). The data has little preference for the value of this parameter within the allowed range, which is not surprising since non-zero synchrotron amplitude is only weakly preferred.

Tightening the prior on the dust/sync correlation parameter from the baseline $-1 < \epsilon < 1$ to $\epsilon = 0$ produces a small downshift in the r constraint curve (black to cyan), as expected given what we already saw in Fig. 15. We show this case as we will invoke it when adding dust decorrelation to the model in Appendix F below. Putting a Gaussian prior on the dust/sync correlation with mean/σ of 0.48/0.50 [7] produces a smaller

TABLE II. Jackknife PTE values from χ^2 and χ (sum of deviations) tests for *Keck Array* 220 GHz data taken in 2015.

Jackknife	Band powers 1–5 χ^2	Band powers 1–9 χ^2	Band powers 1–5 χ	Band powers 1–9 χ
Deck jackknife				
EE	0.515	0.198	0.918	0.365
BB	0.024	0.028	0.008	0.178
EB	0.343	0.551	0.359	0.383
Scan Dir jackknife				
EE	0.962	0.968	0.643	0.579
BB	0.154	0.261	0.579	0.754
EB	0.713	0.896	0.631	0.447
Tag Split jackknife				
EE	0.030	0.014	0.715	0.976
BB	0.327	0.587	0.966	0.948
EB	0.483	0.840	0.234	0.431
Tile jackknife				
EE	0.008	0.026	0.228	0.208
BB	0.242	0.469	0.846	0.850
EB	0.138	0.377	0.597	0.643
Phase jackknife				
EE	0.549	0.858	0.966	0.928
BB	0.343	0.281	0.768	0.479
EB	0.447	0.271	0.669	0.727
Mux Col jackknife				
EE	0.263	0.647	0.257	0.166
BB	0.567	0.693	0.116	0.257
EB	0.936	0.752	0.509	0.719
Alt Deck jackknife				
EE	0.968	0.844	0.573	0.824
BB	0.030	0.172	0.409	0.539
EB	0.517	0.425	0.331	0.106
Mux Row jackknife				
EE	0.695	0.611	0.166	0.094
BB	0.840	0.609	0.649	0.168
EB	0.509	0.311	0.605	0.347
Tile/Deck jackknife				
EE	0.675	0.220	0.768	0.182
BB	0.968	0.990	0.681	0.834
EB	0.972	0.994	0.363	0.246
Focal Plane inner/outer jackknife				
EE	0.020	0.038	0.010	0.016
BB	0.108	0.313	0.032	0.026
EB	0.012	0.040	0.509	0.433
Tile top/bottom jackknife				
EE	0.210	0.108	0.076	0.028
BB	0.030	0.096	0.010	0.006
EB	0.709	0.581	0.685	0.549
Tile inner/outer jackknife				
EE	0.503	0.637	0.503	0.828
BB	0.531	0.549	0.317	0.465
EB	0.477	0.471	0.826	0.723
Moon jackknife				
EE	0.507	0.671	0.910	0.649
BB	0.942	0.894	0.281	0.267
EB	0.639	0.756	0.389	0.539
A/B offset best/worst				
EE	0.561	0.854	0.066	0.082
BB	0.273	0.457	0.443	0.257
EB	0.531	0.569	0.425	0.441

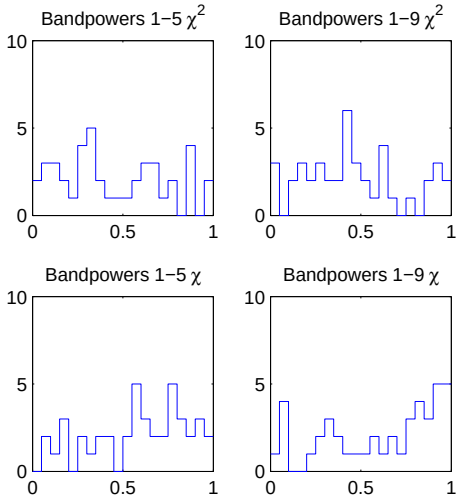

FIG. 10. Distributions of the jackknife χ^2 and χ PTE values for the *Keck Array* 2014 & 2015 95 GHz data over the tests and spectra given in Table I. This figure is analogous to Fig. 12 of Ref. [4].

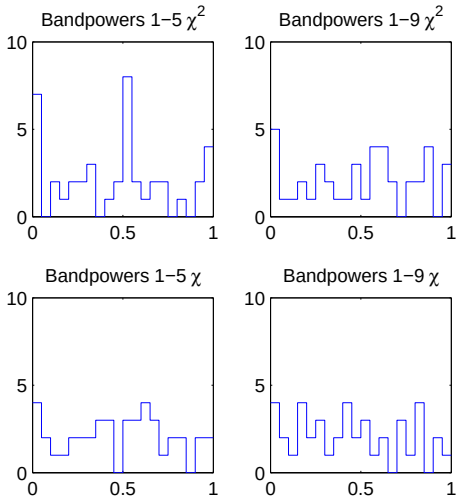

FIG. 11. Distributions of the jackknife χ^2 and χ PTE values for the *Keck Array* 2015 220 GHz data over the tests and spectra given in Table II.

downshift in r than setting $\epsilon = 0$ (comparing yellow and cyan).

We explore the effect of uncertainty in the measured bandpasses for BICEP/*Keck* 95, 150 and 220 GHz channels. We expect such difference to be small and parameterize it as a fractional shift in the band center. We include one parameter for each frequency plus a correlated shift applied to all three channels. For each parameter, we use a Gaussian prior with mean/σ of 0/0.02. These potential bandcenter shifts have little effect on the likelihood (black to dashed blue).

In the baseline analysis, the lensing amplitude is fixed

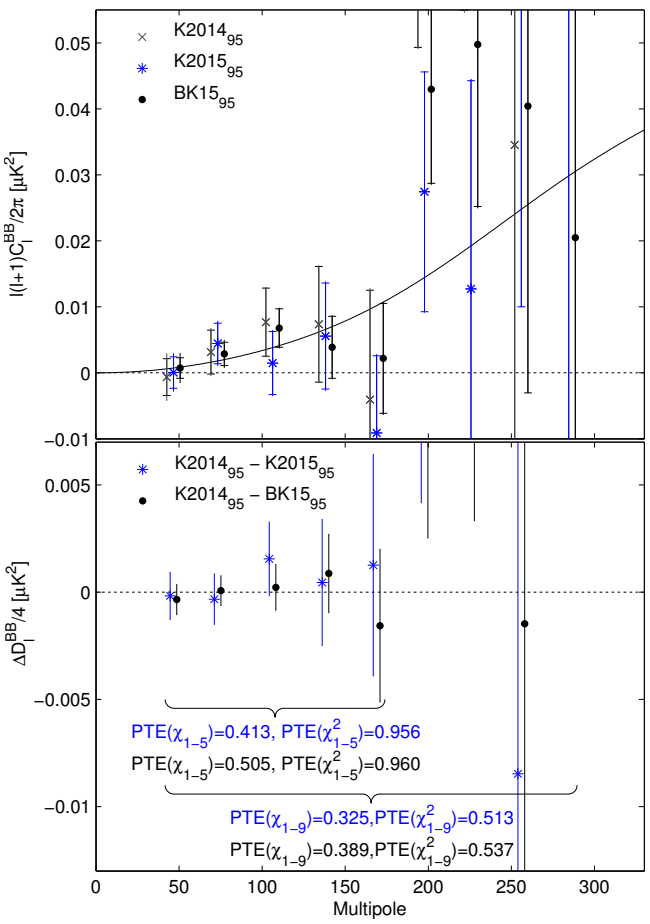

FIG. 12. *Upper:* Comparison of the 95 GHz BB auto-spectrum as previously published (K2014$_{95}$), for 2015 alone (K2015$_{95}$), and for the combination of the two (BK15$_{95}$). The inner error bars are the standard deviation of the lensed-ΛCDM+noise simulations, while the outer error bars include the additional fluctuation induced by the dust signal. Note that neither of these uncertainties are appropriate for comparison of the band power values—for this see the lower panel. (For clarity the sets of points are offset horizontally.) *Lower:* The difference of the pairs of spectra shown in the upper panel divided by a factor of four. The error bars are the standard deviation of the pairwise differences of signal+noise simulations which share common input skies (the simulations used to derive the outer error bars in the upper panel). Comparison of these points with null is an appropriate test of the compatibility of the spectra, and the PTE of χ and χ^2 are shown. This figure is similar to Fig. 13 of Ref. [4].

to the ΛCDM expected value ($A_{\rm L}^{\rm BB} = 1$). Relaxing this assumption we obtain the results shown in Fig. 17. With a unifrom prior, and marginalizing over all other parameters, we obtain $A_{\rm L} = 1.15^{+0.16}_{-0.14}$. The zero-to-peak likelihood ratio is 1.3×10^{-17}, and the probability to have a lower value is 5.8×10^{-19}, which corresponds to a 8.8σ detection. This is the most significant detection of lensing using B-mode polarization to date. Due to the degeneracy between r and $A_{\rm L}$, the r likelihood curve shifts down. If we impose a prior from Planck, $A_{\rm L} = 0.95 \pm 0.04$ [11], the recovered r likelihood curve is almost indistinguishable from the baseline case.

Fig. 18 shows some variations of the data set selection. If we use the BICEP2/*Keck* data only (magenta) the r constraint curve shifts down to peak at zero, while the $A_{\rm d}$ curve broadens slightly, and much larger values of $A_{\rm sync}$ become allowed. Bringing back WMAP (green) produces an even stronger downshift in r, and $A_{\rm sync}$ becomes better constrained. Switching LFI for WMAP (green to yellow) brings r back up a bit and $A_{\rm sync}$ down (note the internal consistency problems of the LFI maps [12]). Adding HFI to BICEP/*Keck*+WMAP (green to red) brings r up and leaves $A_{\rm sync}$ unchanged. BICEP/*Keck*+*Planck* (blue) has almost exactly the same r curve as the baseline but a considerably wider $A_{\rm sync}$ curve. We can understand the behaviors in the $A_{\rm sync}$ curves, at least in part, by noting that in Fig. 13 the BK$_{95}\times$W$_{23}$ bandpowers are positive while the BK$_{95}\times$P$_{30}$ bandpowers are negative.

One additional variation which we explore is to include the EE spectra (and hence also EB) in the fit under the assumption that $EE/BB = 2$ for dust and synchrotron, as is shown to be close to the case in Refs [7] and [13]. (While we have not included the EE jackknife tests in this, or previous, papers they also produce distributions of χ and χ^2 PTE values which are consistent with uniform.) As we can see in Fig. 2, EE spectra such as BK$_{220}\times$P$_{353}$ and P$_{353}\times$P$_{353}$ certainly carry information on the amplitude of the dust emission and can presumably help indirectly to constrain r. In Fig. 18 adding EE results in a small upshift in r and significant tightening of the constraints on $A_{\rm d}$ and $A_{\rm sync}$. We will consider adding EE to the baseline in future analyses, marginalizing over some range in the EE/BB ratios.

At first glance it may appear surprising how large the shifts in the r constraint are under the variations of the data selection shown in Fig. 18, and that many of the shifts are downward. However, when viewing the equivalent plots for the standard lensed-ΛCDM+dust+noise simulation realizations—which contain no tension between the data sets—the qualitative impression in many cases is similar. Note that while we verify in the next section that the baseline r constraint is unbiased, we have not tested this for the data set variations explored here.

3. Likelihood Validation

The interpretation of r likelihood curves such as the one shown in the upper left panel of Fig. 4 is not necessarily straightforward. Since the parameters are restricted to, and marginalized over, physical values only, biases can result. For instance, in a scenario where two parameters are fully degenerate, power will be assigned on average equally between them, and both will be biased low, with the curves for greater than 50% of realizations

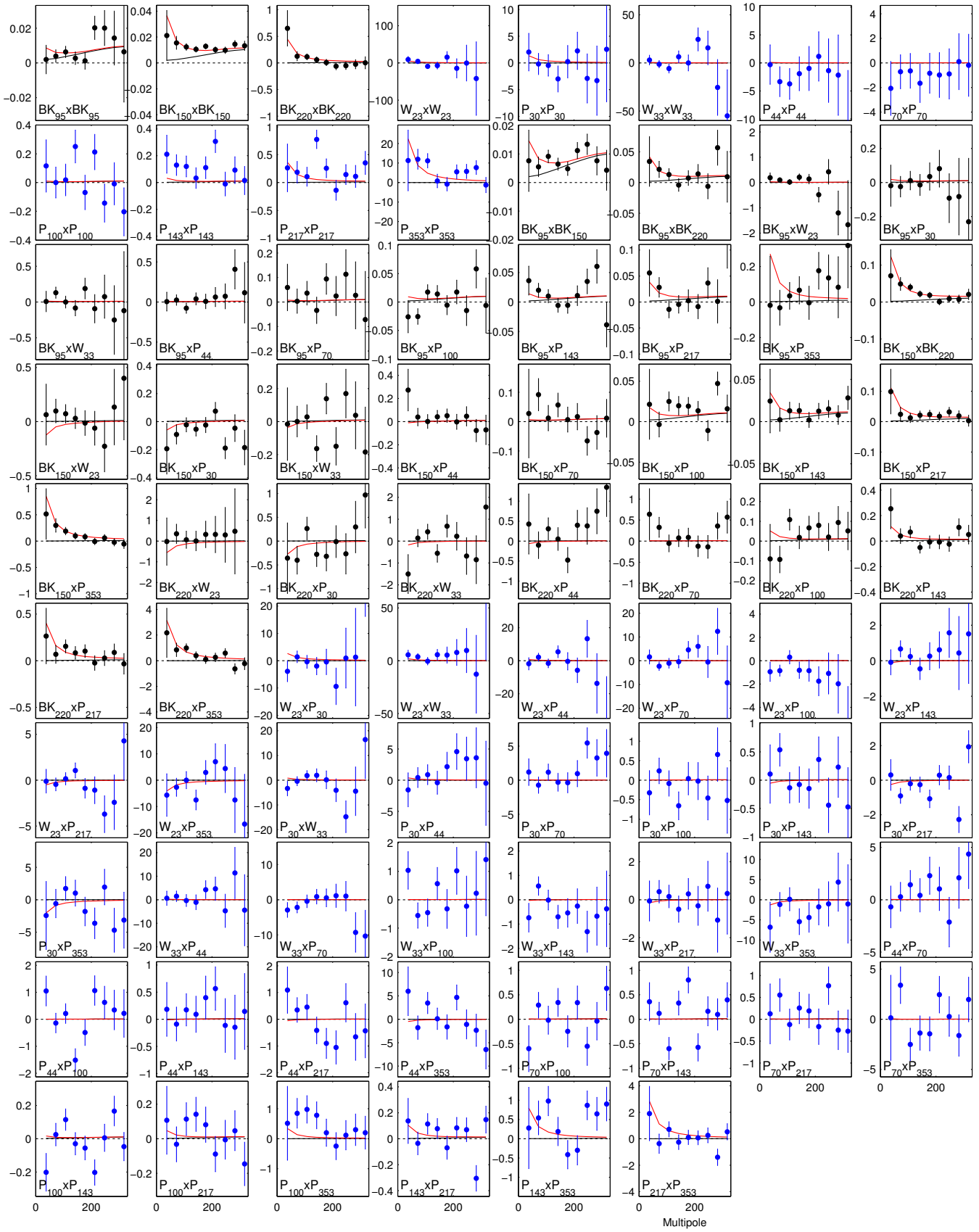

FIG. 13. The full set of BB auto- and cross-spectra from which the joint model likelihood is derived. In all cases the quantity plotted is $100\ell C_\ell/2\pi$ (μK^2). Spectra involving BICEP2/*Keck* data are shown as black points while those using only WMAP/*Planck* data are shown as blue points. The black lines show the expectation values for lensed-ΛCDM, while the red lines show the expectation values of the maximum likelihood lensed-ΛCDM+r+dust+synchrotron model ($r = 0.020$, $A_{\rm d,353} = 4.7\,\mu{\rm K}^2$, $\beta_{\rm d} = 1.6$, $\alpha_{\rm d} = -0.58$, $A_{\rm sync,23} = 1.5\,\mu{\rm K}^2$, $\beta_{\rm s} = -3.0$, $\alpha_{\rm s} = -0.27$, $\epsilon = -0.38$), and the error bars are scaled to that model.

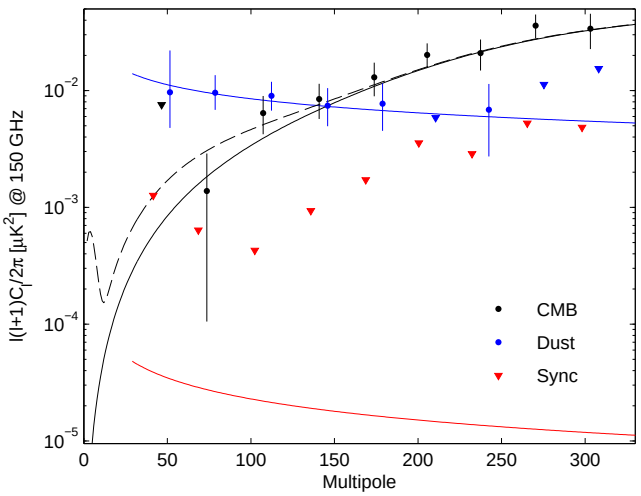

FIG. 14. Spectral decomposition of the BB data into synchrotron (red), CMB (black) and dust (blue) components at 150 GHz. The decomposition is calculated independently in each bandpower, marginalizing over $\beta_{\rm d}$, $\beta_{\rm s}$ and ϵ with the same priors as the baseline analysis. Error bars denote 68% credible intervals, with the point marking the most probable value. If the 68% interval includes zero, we also indicate the 95% upper limit with a downward triangle. (For clarity the sets of points are offset horizontally.) The solid black line shows lensed-ΛCDM with the dashed line adding on top an $r_{0.05} = 0.02$ tensor contribution. The blue/red curves show sync/dust models consistent with the baseline analysis ($A_{\rm d,353} = 4.6\,\mu{\rm K}^2$, $\beta_{\rm d} = 1.6$, $\alpha_{\rm d} = -0.4$ and $A_{\rm sync,23} = 1.0\,\mu{\rm K}^2$, $\beta_{\rm s} = -3.1$, $\alpha_{\rm s} = -0.6$ respectively).

peaking at zero when the true values are zero. To investigate we make full `COSMOMC` runs on the ensemble of lensed-ΛCDM+dust+noise simulations. The left panel of Fig. 19 shows the resulting r constraint curves, while the right panel shows that the CDF of the zero-to-peak likelihood ratios closely follows the simple analytic ansatz $\frac{1}{2}\left(1 - f\left(-2\log L_0/L_{\rm peak}\right)\right)$ where f is the χ^2 CDF (for one degree of freedom). We find that 53% of the simulations peak at zero, and 19% have a lower zero-to-peak ratio than the real data—i.e. show more evidence for r when the true value is in fact zero. This study provides powerful empirical evidence that the real data r constraint curve can be taken at face value, provided the assumed foreground parameterization is an adequate description of reality.

An alternate (and much faster) likelihood validation exercise is to run maximum likelihood searches, with non-physical parameter values allowed (such as negative r). When running on simulations generated according to the model being re-fit, we then have an a priori expectation that the input parameter values should be recovered in the mean. Fig. 20 shows the results when running on the standard lensed-ΛCDM+dust+noise simulations, with the same priors as for the baseline analysis—the input values are recovered in the mean. The first row of Table III summarizes: $\sigma(r) = 0.020$, and bias in the mean value is small as compared to the noise. We prefer this $\sigma(r)$ measure of the intrinsic constraining power of the experiment since it is independent of the particular noise fluctuation that is present in the real data.

4. Exploration of Alternate Foreground Models

We now extend the maximum likelihood validation study to simulations using third-party foreground models. These models do not necessarily conform to our foreground parameterization and therefore when fit to it may potentially produce bias in r at levels relevant compared to the noise. The second and subsequent rows of Table III summarize the results. The third-party models provide only a single realization of the foreground sky, and we add it on top of each of the lensed-ΛCDM+noise realizations that are used in the standard simulations.

TABLE III. Uncertainty and bias on r in simulations using Gaussian and 3rd party foreground models. (The numbers in parentheses suffer from disagreement between the priors and the model so bias is expected—see text for details.)

	$\overline{A_d}$	$\overline{A_s}$	$\sigma(r)$, $\overline{r}/\sigma(r)$		
Model	(μK^2)	(μK^2)	$\beta_{\rm d}$ prior	$\beta_{\rm d}$ free	with decorr.
Gaussian	3.8	0.1	0.020, $+0.1\sigma$	0.023, 0.0σ	0.021, $+0.0\sigma$
PySM 1	10.9	1.1	0.026, $+0.2\sigma$	0.028, $+0.2\sigma$	0.028, $+0.1\sigma$
PySM 2	24.2	0.9	0.028, $+0.1\sigma$	0.029, $+0.1\sigma$	0.032, $+0.1\sigma$
PySM 3	12.1	1.1	(0.030, $+0.4\sigma$)	0.031, $+0.1\sigma$	(0.032, $+0.2\sigma$)
MHDv2	2.9	5.6	0.020, $+0.2\sigma$	0.027, -0.2σ	0.021, -0.1σ
G. Decorr.	4.6	0.1	(0.023, $+1.5\sigma$)	(0.026, $+1.3\sigma$)	0.022, $+0.0\sigma$

The PySM models 1, 2 and 3 are a1d1f1s1, a2d4f1s3 and a2d7f1s3 respectively, with the letters indicating AME (a), dust (d), free-free (f) & synchrotron (s), and the numbers referring to the various models of each as described in the PySM paper [14]. The a1 and f1 models are unpolarized and hence not relevant. The a2 model uses a *Planck* `Commander` [15] derived template and (dust) polarization angles together with a conservative 2% polarization fraction. No account for AME is made in our parametric model so this could potentially result in bias. The d1 model again uses *Planck* `Commander` derived templates for both the 353 GHz Q/U patterns and the T_d and $\beta_{\rm d}$ spectral parameters. The dust SED thus varies spatially, and this model therefore implements decorrelation of the dust pattern at some level (which in practice is found to be very small). Model d4 generalizes model d1 to the two temperature FDS model [16]. Model d7 is a sophisticated physical model of dust grains as described in Ref. [17] which does not necessarily conform to the modified blackbody SED. The s1 model takes the WMAP 23 GHz Q/U maps and rescales them according to a power law using a spectral index map, and the s3 model adds on top of this a (spatially uniform) curvature of the synchrotron SED. The WMAP and *Planck*

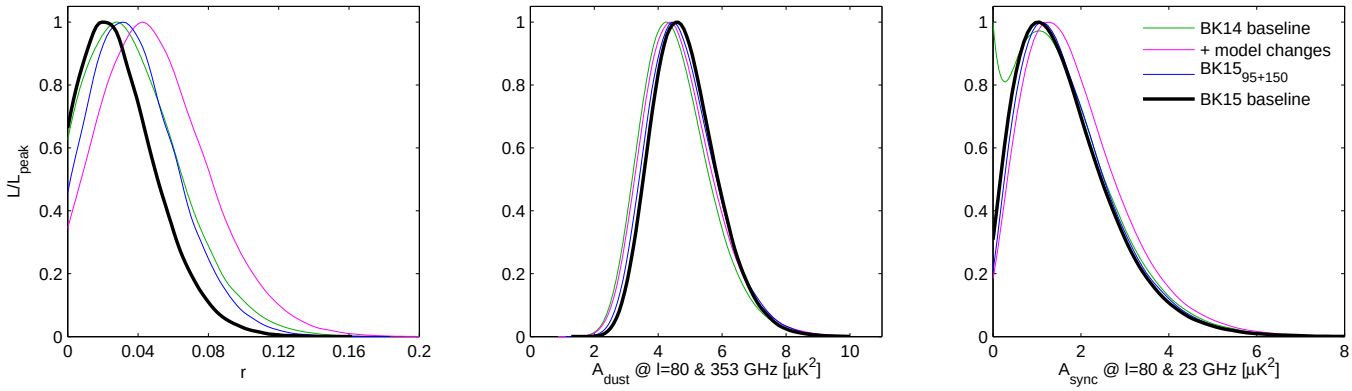

FIG. 15. Evolution of the BK14 analysis to the "baseline" analysis as defined in this paper—see Appendix E 1 for details.

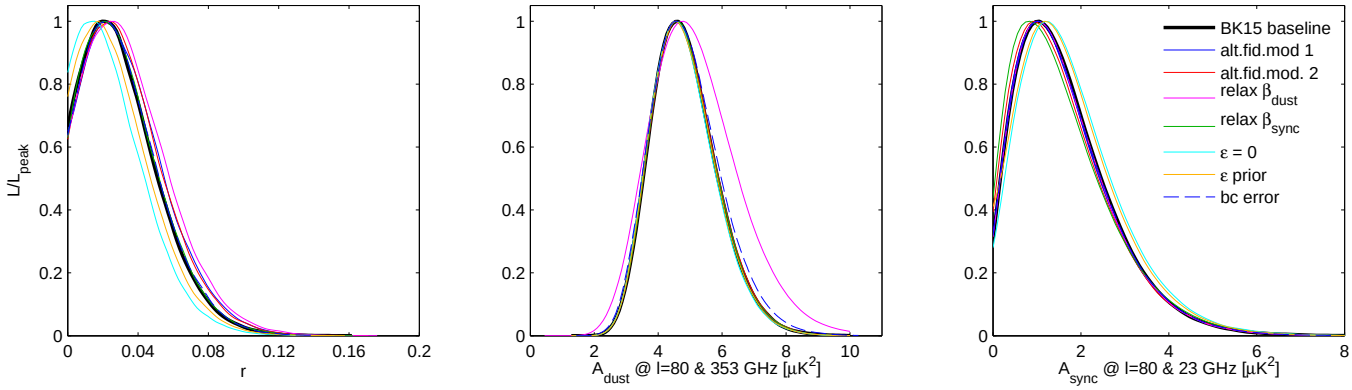

FIG. 16. Likelihood results when varying the analysis choices—see Appendix E 2 for details.

polarization templates are all noise dominated at smaller angular scales, so PySM filters out this noise and fills back in Gaussian realizations of foreground structure according to the recipe described in Sec. 3.1 of the PySM paper [14].

We see in Table III that the PySM models predict considerably higher dust power in the BICEP/*Keck* field than is actually observed and that this pushes up $\sigma(r)$ somewhat as compared to the Gaussian results. The dust amplitude is sufficiently high in these models that β_d becomes well constrained for the noise levels and frequency range of the BK15 data—the prior on β_d can therefore be relaxed, and this is actually necessary for the PySM 3 model where the value of β_d preferred by the model is outside of the prior range, and bias on r results if the prior is not relaxed.

The model labeled "MHDv2" is based on simulations of the Galactic magnetic field [18] and naturally produces correlated dust and synchrotron emission. Since this model contains no explicit experimental data there is no noise issue, and the generated structure is non-Gaussian across the full range of ℓ. This model gives a higher level of synchrotron than that which is preferred by the BICEP/*Keck* data ($A_{\mathrm{sync},23} = 5.6\,\mu\mathrm{K}^2$ as compared to the maximum likelihood value of $1.5\,\mu\mathrm{K}^2$ and 95% upper limit of $A_{\mathrm{sync},23} < 3.7\,\mu\mathrm{K}^2$). This model also produces bias in the mean value of r that is small compared to the noise level.

We conclude that none of the considered models produces relevant bias on r when fitted to our foreground parameterization for the current experimental noise levels. These models span a variety of assumptions and methods and in some cases predict levels of foreground contamination much stronger than we actually observe in our field. However, there is no guarantee that the real foregrounds do not in fact produce greater bias than any of the considered models. We note that all of the above models produce dust decorrelation that is negligibly small compared to the current noise levels.

Appendix F: Adding dust decorrelation

The simplest possible model of a given component of the polarized foreground emission (e.g. dust or synchrotron) is that it presents a fixed spatial pattern on

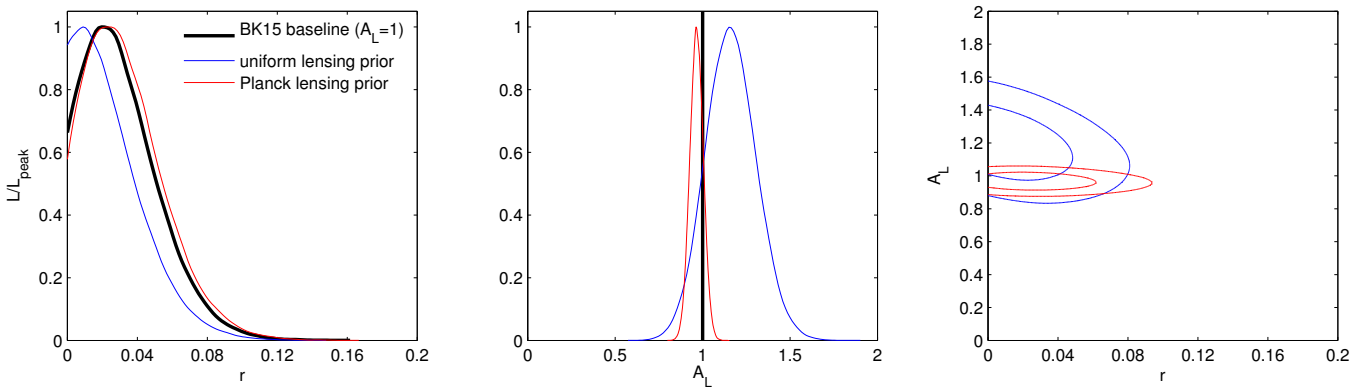

FIG. 17. Likelihood results when allowing the lensing amplitude to be a free parameter—see Appendix E 2 for details.

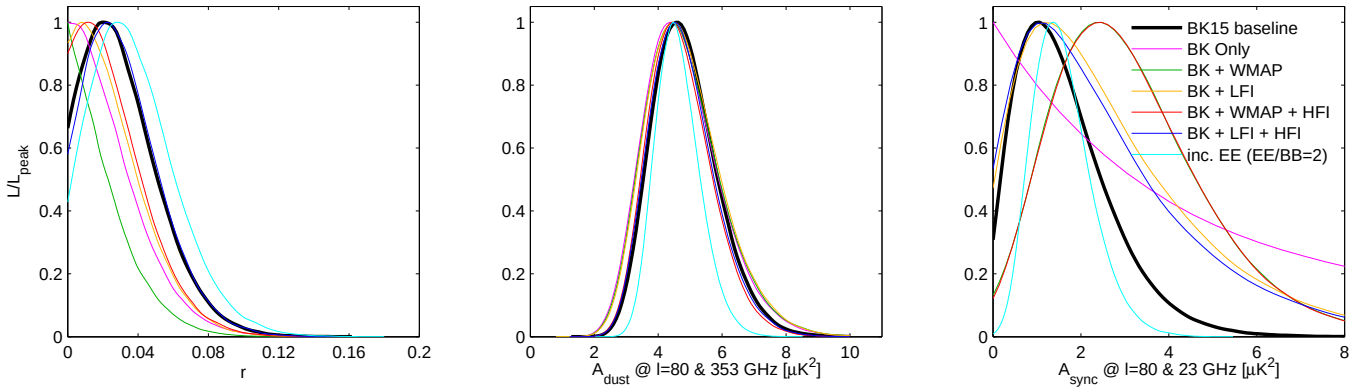

FIG. 18. Likelihood results when varying the data set selection—see Appendix E 2 for details.

the sky which scales with frequency according to a single SED. In this case the cross-spectrum between any two given frequencies is simply the geometric mean of the respective auto-spectra. In reality the morphology of the polarization pattern will inevitably vary as a function of frequency at some level. If Q and U at each given point on the sky deviate in sympathy away from the mean SED then the polarized intensity map will evolve as a function of frequency, but the polarization angles will remain constant. If Q and U deviate independently from the mean SED then both polarization intensity and angle will be functions of observing frequency. In either case the cross-spectra will be suppressed with respect to the geometric mean of the auto-spectra—a phenomenon which we refer to as decorrelation.

Planck Intermediate Paper XXX [19] looked for suppression of the cross-spectral amplitudes in Figs. 6 & E.1 and did not find any evidence for decorrelation. However, that analysis was implicitly weighted towards lower ℓ. Later *Planck* Intermediate Paper L [20, hereafter PIPL] examined the cross-spectrum between 220 & 353 GHz as a function of ℓ and found evidence for a suppression effect which increased with ℓ and also when going to cleaner regions of sky (as determined by neutral hydrogen column density—see Fig. 3 of that paper). More recently, Ref. [21] re-analyzed the now public *Planck* data and found no evidence for a detection of dust decorrelation. Finally the *Planck* team revisited the issue again in *Planck* Intermediate Paper LIV [7, hereafter PIPLIV] and this time state that "We find no evidence for a loss of correlation."

Decorrelation certainly exists at some level—the question is whether that level is relevant as compared to the current instrumental noise. To search for evidence of decorrelation in the BK15 data we add decorrelation of the dust pattern to our parametric model. We define the correlation ratio of the dust

$$\Delta_{\rm d} = \frac{\mathcal{D}_{80}(217\times353)}{\sqrt{\mathcal{D}_{80}(217\times217)\mathcal{D}_{80}(353\times353)}}, \qquad \text{(F1)}$$

where $\mathcal{D}_{80}$ is the dust power at $\ell = 80$. This makes $\Delta_{\rm d}$ close to equivalent to $\mathcal{R}_{80}^{BB}$ as defined by PIPL and PIPLIV. We scale to other frequency combinations using the factor

$$f(\nu_1,\nu_2) = \frac{(\log(\nu_1/\nu_2))^2}{(\log(217/353))^2}, \qquad \text{(F2)}$$

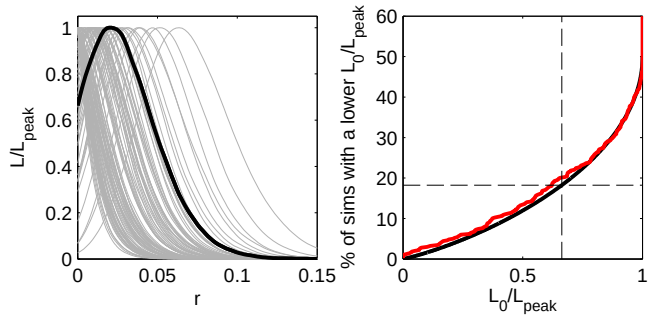

FIG. 19. *Left:* Likelihood curves for r when running the baseline analysis on each of the lensed-ΛCDM+dust+noise simulations—half of them peak at zero. The real data curve is shown overplotted in heavy black. *Right:* The CDF of the zero-to-peak ratio (red) of the curves shown at right as compared to the simple analytic ansatz (solid black) $\frac{1}{2}\left(1 - f\left(-2\log L_0/L_{\rm peak}\right)\right)$ where f is the χ^2 CDF (for one degree of freedom). About one fifth of the simulations offer more evidence for non-zero r than the real data when the true value is actually zero (dashed black).

as suggested by PIPL.

Fig. 2 of PIPL suggests that decorrelation grows with increasing ℓ, although in Sec. 4 they assume flat with ℓ. In this paper we consider two possible scalings

$$g(\ell) = \begin{cases} 1 & \text{flat case} \\ (\ell/80) & \text{linear case} \end{cases} . \qquad \text{(F3)}$$

Since the ℓ range we are concerned with is not broad this choice turns out to make little practical difference.

The above scalings can produce extreme, and non-physical, behavior for widely separated frequencies and low/high ℓ. We therefore re-map the nominal value using the following function

$$\Delta'_{\rm d}(\nu_1, \nu_2, \ell) = \exp\left[\log(\Delta_{\rm d})\, f(\nu_1, \nu_2)\, g(\ell)\right], \qquad \text{(F4)}$$

such that $\Delta'_{\rm d}$ remains in the range 0 to 1 for all values of f and g. We note that for the frequency scaling this becomes the same as Eqn. 14 of Ref. [22] which is shown in that paper to correspond to a Gaussian spatial variation in the foreground spectral index. (This is also used in PIPLIV.) For the moment we defer consideration of models which have both decorrelation of the dust pattern and correlation of the dust and synchrotron patterns simultaneously, setting $\epsilon = 0$ whenever we allow $\Delta_{\rm d} \neq 1$. Note that in Fig. 16 we see that setting $\epsilon = 0$ produces only small changes from the baseline analysis.

Fig. 2 shows the power spectra of the frequency bands which have the most power to constrain the dust contribution to the model. We can see visually that the (non-decorrelated) model from our previous BK14 analysis which is plotted there appears to be a good explanation of the observations (and in Appendix D it was shown formally that the new BK15 maximum likelihood model is compatible with the data). PIPL states that the mean neutral hydrogen column density in the BICEP2/*Keck* field is $\sim 1.6 \times 10^{20}\,{\rm cm}^{-2}$ for which their Eqn. 6 gives a predicted correlation ratio value $\mathcal{R}^{BB}_{50-160}(217, 353) = 0.83$. To illustrate the effect of decorrelation in Fig. 2 we also re-plot the BK14 model modified with $\Delta_{\rm d} = 0.85$ as the dashed red lines—this leaves the auto-spectra unchanged while suppressing the cross-spectra. The 150×353 data appears to weakly disfavor the change while the 95×353 weakly favors it. The above is simply for the purposes of illustration—we proceed below to include decorrelation and re-fit the model.

We expand the baseline likelihood analysis to include decorrelation and show results in Fig. 21. We consider several choices of prior on the $\Delta_{\rm d}$ parameter: i) Based on Table 1 of Ref. [21] and Table 3 of Ref. [7] we set a Gaussian prior with mean/σ of 0.95/0.05 (truncated above 1), flat with ℓ. ii) A Gaussian prior with mean/σ of 1.00/0.05, linear with ℓ. iii) A uniform prior 0 to 1, linear with ℓ. All of these choices result in the r likelihood curve shifting down and peaking at zero. However, note that introducing $\Delta_{\rm d}$ in a likelihood analysis which marginalizes only over the physically meaningful range $\Delta_{\rm d} \leq 1$ can result in a downward bias on r even in the absence of a real decorrelation effect. For a given set of bandpowers it is possible to explain observed power in cross-spectra such as 150×353 with a higher value of A_d in combination with a lower value of $\Delta_{\rm d}$. The auto-spectra resist this preventing strong degeneracy, but a net bias still results. When we repeat the exercise of Fig. 19 running the full analysis on the standard lensed-ΛCDM+dust+noise simulations (which do not contain decorrelation), but include the decorrelation parameter in the analysis, we find that 72% of the r curves peak at zero, and many of the $\Delta_{\rm d}$ curves peak below 1. We therefore choose not to include the decorrelation parameter in our baseline analysis at this time.

To check that the machinery remains unbiased when running maximum likelihood searches we repeat the exercise of Appendix E 3 but this time including the decorrelation parameter $\Delta_{\rm d}$ and allowing it to take values greater than one. To do this in a symmetrical manner we use

$$\Delta'_{\rm d}(\nu_1, \nu_2, \ell) = 2 - \exp\left[\log(2 - \Delta_{\rm d})\, f(\nu_0, \nu_1)\, g(\ell)\right]. \qquad \text{(F5)}$$

In this exercise we take the linear ℓ scaling. Fig. 22 shows the results for the standard lensed-ΛCDM+dust+noise simulations which contain no decorrelation. We see that $\Delta_{\rm d} = 1$ is recovered, and r remains unbiased. We also show results for a toy highly decorrelated model which uses $\Delta_{\rm d} = 0.85$ and linear scaling with ℓ, following Eqns. F2–F4. The input parameters of this model are also recovered in the mean. Finally we run the analysis with decorrelation on the third-party foreground models and give results for all the models in Table III. As expected we see that the decorrelated simulations produce bias when re-analyzed without allowing decorrelation in the model.

Running a maximum likelihood search including decorrelation on the real data we obtain $r_{0.05} = -0.012$,

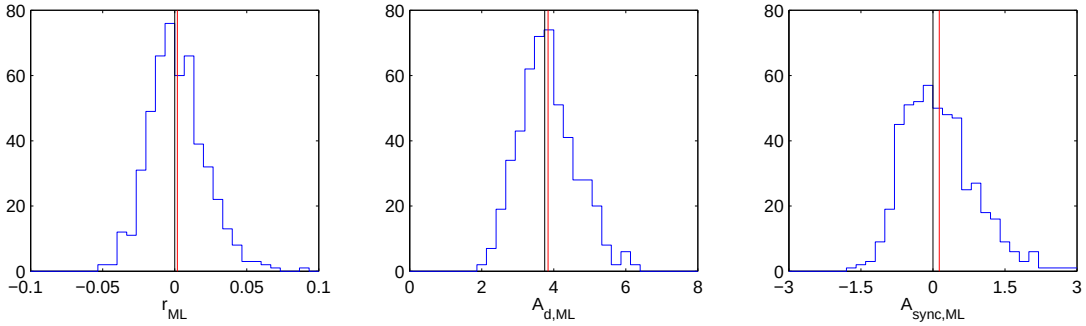

FIG. 20. Results of a validation test running maximum likelihood search on simulations of a lensed-ΛCDM+dust+noise model with no synchrotron ($A_{\rm d,353} = 3.75\,\mu{\rm K}^2$, $\beta_{\rm d} = 1.6$, $\alpha_{\rm d} = -0.4$, $A_{\rm sync} = 0$). The baseline priors are applied on $\beta_{\rm d}$, $\beta_{\rm s}$, $\alpha_{\rm d}$, $\alpha_{\rm s}$ and ϵ. The blue histograms are the recovered maximum likelihood values with the red lines marking their means and the black lines showing the input values. In the left panel $\sigma(r) = 0.020$. See Appendix E 3 for details.

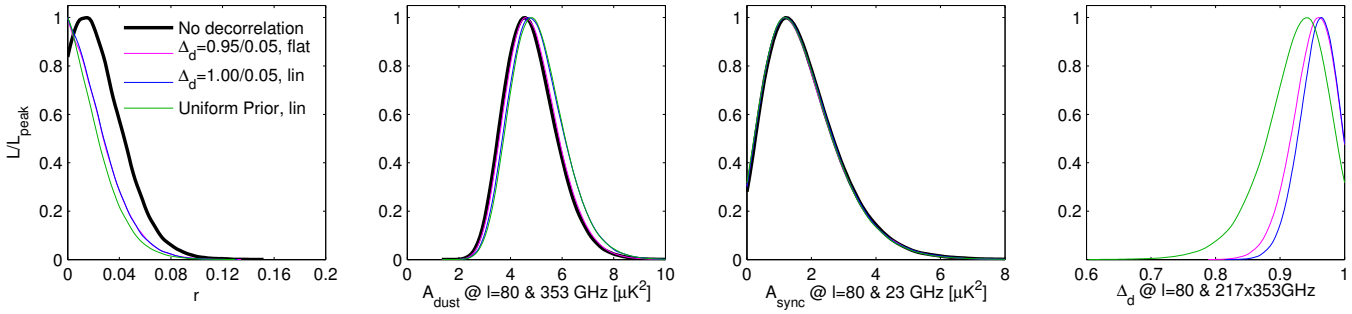

FIG. 21. Likelihood results when allowing dust decorrelation—see Appendix F for details.

$A_{\rm d,353} = 5.0\,\mu{\rm K}^2$, $A_{\rm sync,23} = 1.4\,\mu{\rm K}^2$, $\beta_{\rm d} = 1.6$, $\beta_{\rm s} = -3.1$, $\alpha_{\rm d} = -0.38$, $\alpha_{\rm s} = -0.52$, and $\Delta_{\rm d} = 0.92$, i.e. $A_{\rm d}$ shifts up a bit, r shifts down a bit, and $\Delta_{\rm d}$ is a little less than one. This model has a χ^2 versus the data of 759 to be compared to the baseline model value of 760—the data shows little evidence for decorrelation of the dust pattern. As the data improves in the future the ability to constrain decorrelation while remaining unbiased on r will improve.

Appendix G: Definition of Multicomponent Model

The likelihood analysis uses a parametrized model to describe the bandpower expectation values as a combination of cosmological and foreground signals. The form of this model remains unchanged from BKP and BK14 except for the addition of foreground decorrelation (described in Appendix F). However, the choice of free parameters and priors has evolved over time due to improved BICEP/*Keck* data and new information from external sources. The previous papers describe the important features of the model but do not include a complete mathematical formulation, which we provide here.

Equation G1 describes contributions to the BB cross-spectrum between maps at frequencies ν_1 and ν_2 (or auto-spectrum, if $\nu_1 = \nu_2$) from dust, synchrotron, and the spatially-correlated component of dust and synchrotron. Parameter $A_{\rm d}$ specifies the dust power in units of $\mu{\rm K}^2_{\rm CMB}$ at pivot frequency 353 GHz and angular scale $\ell = 80$. Parameter $A_{\rm sync}$ specifies synchrotron power similarly, except with a pivot frequency of 23 GHz. The dust and synchrotron components scale as power laws in ℓ with slopes $\alpha_{\rm d}$ and $\alpha_{\rm s}$, respectively. Note that we define parameters $\alpha_{\rm d}$ and $\alpha_{\rm s}$ as the ℓ scaling of $\mathcal{D}_\ell \equiv \ell\,(\ell+1)\,C_\ell/2\pi$, not C_ℓ .

The level of spatial correlation between dust and synchrotron is set by parameter ϵ. The correlated component scales with ℓ with a slope that is the average of $\alpha_{\rm d}$ and $\alpha_{\rm s}$, meaning that the correlation coefficient is assumed to be constant across all ℓ.

Parameter $\Delta'_{\rm d}$ accounts for decorrelation of the dust pattern between ν_1 and ν_2 and is defined in equation F4. Note that if $\nu_1 = \nu_2$, then $\Delta'_{\rm d} = 1$ (perfect correlation). Parameter $\Delta'_{\rm s}$ describes decorrelation of the synchrotron pattern but is not currently used. We currently do not include foreground decorrelation parameters in the dust–synchrotron correlated component. A complete foreground model would include the full set of correlations between dust and synchrotron fields at ν_1 and the

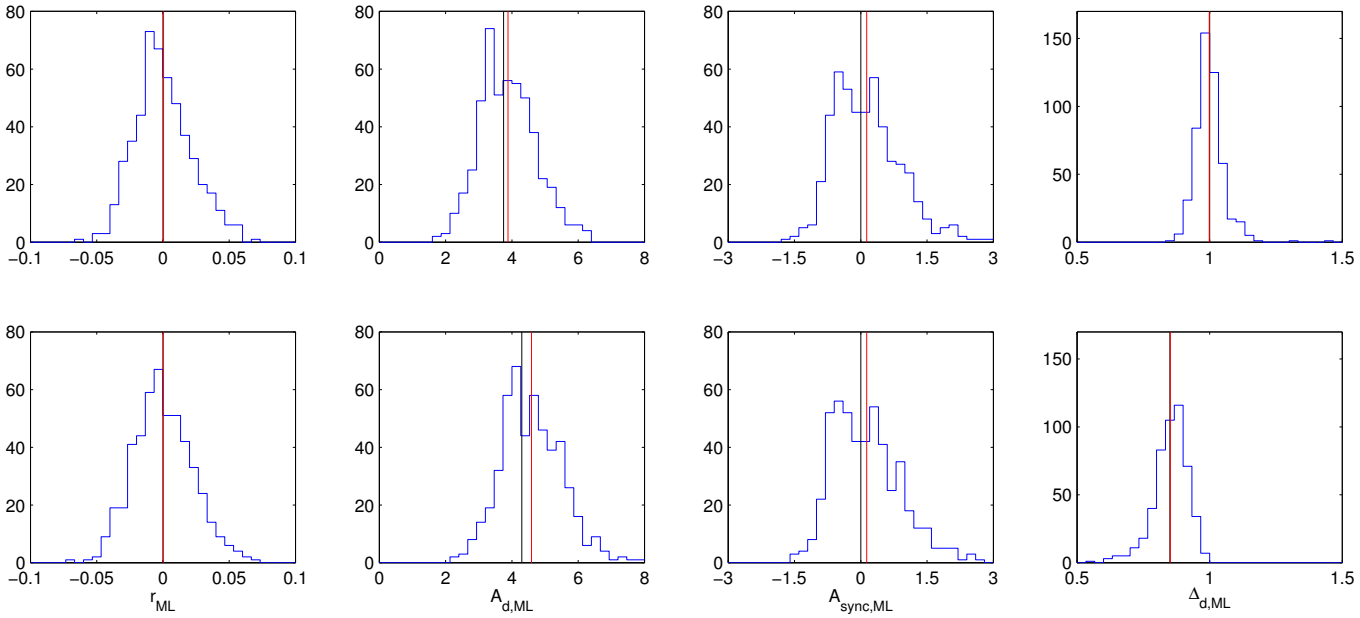

FIG. 22. Validation tests running the likelihood with the dust decorrelation parameter $\Delta_{\rm d}$ included. *Upper row:* Results for the same lensed-ΛCDM+dust+noise simulations shown in Fig. 20. *Lower row:* Results for the toy highly decorrelated dust model. The blue histograms are the recovered maximum likelihood values with the red lines marking their means and the black lines showing the input values. See Appendix F for details.

dust and synchrotron fields at ν_2, but current data offer no guidance about the form of these correlations. For the time being, we consider dust decorrelation only as an extension to models with $\epsilon = 0$, as noted in Appendix F.

Additional coefficients $f_{\rm d}$ and $f_{\rm s}$ capture the scaling of dust and synchrotron power from the pivot frequencies to the actual bandpasses of the maps labeled ν_1 and ν_2. This scaling includes the foreground SED as well as the conversion between $\mu{\rm K}_{\rm CMB}$ units at the pivot frequency and at the target map bandpass. The SED model used for dust is a blackbody with temperature $T_{\rm d} = 19.6$K multiplied by a power law with emissivity spectral index $\beta_{\rm d}$ [8]. The SED model used for synchrotron is a power law with spectral index $\beta_{\rm s}$ defined relative to a Rayleigh-Jeans spectrum. When integrating the SED and unit conversion factors over a map bandpass it is necessary to choose a bandpass convention. We define our bandpass functions to be proportional to the response as a function of frequency to a beam-filling source with uniform spectral radiance (the same convention as used by *Planck* [23]).

$$\mathcal{D}^{\nu_1\times\nu_2}_{\ell,BB} = A_{\rm d}\Delta'_{\rm d} f_{\rm d}^{\nu_1} f_{\rm d}^{\nu_2} \left(\frac{\ell}{80}\right)^{\alpha_{\rm d}} + A_{\rm sync}\Delta'_{\rm s} f_{\rm s}^{\nu_1} f_{\rm s}^{\nu_2} \left(\frac{\ell}{80}\right)^{\alpha_{\rm s}} + \epsilon\sqrt{A_{\rm d}A_{\rm sync}}(f_{\rm d}^{\nu_1} f_{\rm s}^{\nu_2} + f_{\rm s}^{\nu_1} f_{\rm d}^{\nu_2}) \left(\frac{\ell}{80}\right)^{(\alpha_{\rm d}+\alpha_{\rm s})/2} \quad \text{(G1)}$$

The foreground contribution to EE is similar, except that $A_{\rm d}$ and $A_{\rm sync}$ are scaled by the EE/BB ratios for dust and synchrotron, respectively, which are both assumed to be equal to 2 [7, 13]. The model for the EB spectrum is zero, since neither CMB nor foreground signals are expected to break parity symmetry. We do not model the unpolarized foregrounds, nor include $TT/TE/TB$ spectra in the likelihood analysis.

Appendix H: Summary of Simulations

We interpret the single realization of real data through comparison to several sets of simulations. With the exception of the alternate foreground models mentioned in Appendix E 4 above these have all been described and used in our previous papers [1, 4, 9].

We start by generating 499 pseudosimulations of noise by the sign-flip technique [1, 24]. During the addition of multiple data subsets to form the final map we randomly flip the signs to cancel out sky signal. Each sequence is

constructed to have equal weight in positives and negatives, and since the sequences are $> 10^4$ in length the resulting noise realizations are found empirically to be uncorrelated. The mean spectra of these noise simulations are used to debias the real spectra (this being very important for the auto-spectra).

We also generate 499 realizations of lensed and unlensed ΛCDM by resampling timestream from simulated input maps and passing it through the full analysis pipeline (including filtering etc.) [1]. The unlensed simulations are useful to empirically determine the purity delivered by the matrix purification algorithm which is used to extract the B-mode signal in the presence of a much stronger E-mode.

From the simulated signal-cross-signal, noise-cross-noise and signal-cross-noise spectra we can construct the bandpower covariance matrix appropriate for any model containing a set of signal components with given SEDs [9]. When we do this we set to zero any term which has an expectation value of zero (under the assumption that signal and noise are uncorrelated) to reduce the Monte Carlo error in the resulting covariance matrix given the relatively modest number of realizations. We also set to zero the covariance between bandpowers that are separated by more than one bin in ℓ, but, importantly, preserve the covariance between the the auto- and cross-spectra of the different frequency bands. This covariance matrix construction is used for the HL likelihood, and also to provide bandpower uncertainties shown, for example, in Fig. 13.

We also explicitly simulate simple dust input maps as power-law Gaussian realizations (with amplitude set to the observed dust amplitude in the BICEP/*Keck* patch) and pass these through the timestream sampling and pipeline re-mapping operation. They are then added to the lensed-LCDM and noise maps, and taken through to power spectra. We use these when it is important to match the fluctuations present in the real data in detail. One example is in the spectral stability tests shown in Fig. 12. Another example is when determining the PTE of the real data χ^2 value in Appendix D.

Appendix I: Lensing analysis

In Ref. [25], we showed a detection of the gravitational lensing signal using the BK14 E- and B-modes at 150 GHz. We showed that the lensing signal is consistent with the standard ΛCDM model, and the BK14 B-mode spectrum at intermediate scales is dominated by lensing.

At 150 GHz, the sensitivity of BK15 to lensing is almost the same as that of BK14. Reconstructed lensing maps at 95 GHz and 220 GHz are still noisy. However, reconstructing lensing signals from BK15 data is important to test consistency of the data and simulation.

We reconstruct the lensing maps using BK15 data at 95 GHz, 150 GHz and 220 GHz based on the method described in Ref. [25]. Because the *Planck* lensing map has higher signal-to-noise than our reconstructed lensing maps, the BK15 lensing maps are then cross-correlated with the *Planck* lensing map provided by Ref. [26]. Fig. 23 shows the cross correlation of the reconstructed lensing signals between *Planck* and BK15 at each frequency. The amplitudes of the observed lensing spectra relative to the simulated spectra are found to be $A_{\rm L}^{\phi\phi} = 1.24 \pm 0.39$ (95 GHz), 1.14 ± 0.20 (150 GHz) and -1.13 ± 1.87 (220 GHz), respectively. The data are consistent with our baseline simulation, and no spurious behavior is found in the lensing analysis.

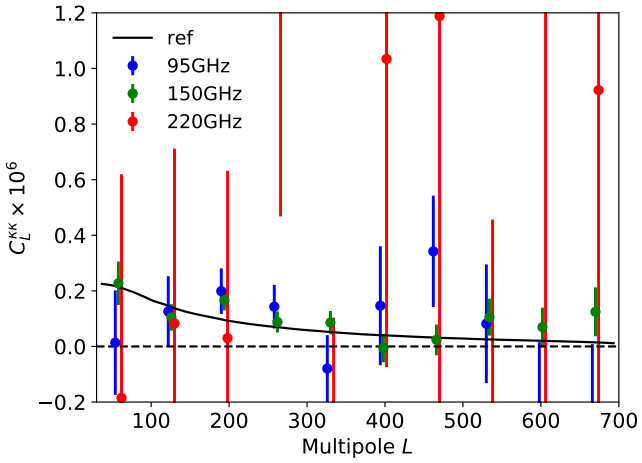

FIG. 23. Cross-correlation of the lensing reconstructions between *Planck* and BK15. We show the spectra for reconstruction using the BK15 95 GHz, 150 GHz and 220 GHz bands. The black solid line shows the theoretical lensing power spectrum.

[1] BICEP2 Collaboration I, Physical Review Letters **112**, 241101 (2014), arXiv:1403.3985

[2] *Keck Array* and BICEP2 Collaborations XI, (to be published)

[3] *Keck Array* and BICEP2 Collaborations V, Astrophys. J. **811**, 126 (2015), arXiv:1502.00643

[4] *Keck Array* and BICEP2 Collaborations VI, Physical Review Letters **116**, 031302 (2016), arXiv:1510.09217

[5] S. Hamimeche and A. Lewis, Phys. Rev. D **77**, 103013 (2008), arXiv:0801.0554

[6] S. K. Choi and L. A. Page, J. Cosmol. Astropart. Phys. **12**, 020 (2015), arXiv:1509.05934

[7] Planck Collaboration 2018 XI, ArXiv e-prints (2018), arXiv:1801.04945

[8] Planck Collaboration Int. XXII, Astron. Astrophys. **576**, A107 (2015), arXiv:1405.0874

[9] BICEP2/*Keck* and *Planck* Collaborations, Physical Review Letters **114**, 101301 (2015), arXiv:1502.00612

[10] U. Fuskeland, I. K. Wehus, H. K. Eriksen, and S. K. Næss, Astrophys. J. **790**, 104 (2014), arXiv:1404.5323

[11] Planck Collaboration 2015 XIII, Astron. Astrophys. **594**, A13 (2016), arXiv:1502.01589

[12] Planck Collaboration 2015 II, Astron. Astrophys. **594**, A2 (2016), arXiv:1502.01583 [astro-ph.IM]

[13] N. Krachmalnicoff, E. Carretti, C. Baccigalupi, G. Bernardi, S. Brown, B. M. Gaensler, M. Haverkorn, M. Kesteven, F. Perrotta, S. Poppi, and L. Staveley-Smith, ArXiv e-prints (2018), arXiv:1802.01145

[14] B. Thorne, J. Dunkley, D. Alonso, and S. Næss, Mon. Not. R. Astron. Soc. **469**, 2821 (2017), arXiv:1608.02841

[15] Planck Collaboration 2015 X, Astron. Astrophys. **594**, A10 (2016), arXiv:1502.01588

[16] D. P. Finkbeiner, M. Davis, and D. J. Schlegel, Astrophys. J. **524**, 867 (1999), astro-ph/9905128

[17] B. Hensley, *On the nature of interstellar grains*, Ph.D. thesis, Princeton University (2015)

[18] A. G. Kritsuk, S. D. Ustyugov, and M. L. Norman, New Journal of Physics **19**, 065003 (2017), arXiv:1705.01912

[19] Planck Collaboration Int. XXX, Astron. Astrophys. **586**, A133 (2016), arXiv:1409.5738

[20] Planck Collaboration Int. L, Astron. Astrophys. **599**, A51 (2017), arXiv:1606.07335

[21] C. Sheehy and A. Slosar, Phys. Rev. D **97**, 043522 (2018), arXiv:1709.09729

[22] F. Vansyngel, F. Boulanger, T. Ghosh, B. D. Wandelt, J. Aumont, A. Bracco, F. Levrier, P. G. Martin, and L. Montier, Astron. Astrophys. **603**, A62 (2017), arXiv:1611.02577

[23] Planck Collaboration IX, Astron. Astrophys. **571**, A9 (2014)

[24] A. van Engelen, R. Keisler, O. Zahn, K. A. Aird, B. A. Benson, L. E. Bleem, J. E. Carlstrom, C. L. Chang, H. M. Cho, T. M. Crawford, A. T. Crites, T. de Haan, M. A. Dobbs, J. Dudley, E. M. George, N. W. Halverson, G. P. Holder, W. L. Holzapfel, S. Hoover, Z. Hou, *et al.*, Astrophys. J. **756**, 142 (2012), arXiv:1202.0546

[25] *Keck Array* and BICEP2 Collaborations VIII, Astrophys. J. **833**, 228 (2016), arXiv:1606.01968

[26] P. A. R. Ade *et al.* (Planck), Astron. Astrophys. **594**, A15 (2016), arXiv:1502.01591